FORMULAS FOR CALCULATING NUMBER OF FRUITS REQUIRED FOR ADEQUATE SAMPLE FOR ANALYSIS[1]

F. E. DENNY

When taking samples of variable fruits, as oranges for example, it is important to obtain an approximation of the number of fruits that should be included in the sample, in order that the results of the analyses shall be sufficiently accurate for the purpose of the investigation. It is the object of this paper to give formulas which may be used in such cases; to illustrate their use by numerical examples; to indicate the reliability that may be placed upon them; and to show the results that were obtained in applying them to the analysis of citrus fruits.

The first step consisted in obtaining a measure of the variability of the fruit in question. In the case of citrus, this was accomplished by analyzing individual fruits, since one fruit was found to yield enough material for the analytical work performed. If smaller fruits, such as plums, were used, it would be necessary to increase the sample to half a dozen, or a dozen, or some other number that would make a convenient sample with which to work, but the results of the analysis of each of the chosen units should be tabulated separately. From these data the probable error of a single sample was found, and this value formed the starting point for the calculations made in formulas described in later paragraphs.

Variability in composition of individual oranges in single sample

Fifty-one oranges were taken at random from a single tree. These fruits were all of good marketable quality, and were apparently free from diseases, insect injuries, and bruises. They were uniform in color, but of course variable in size. The fruits were analyzed individually and the results for each fruit tabulated

[1] Published by permission of the Secretary of Agriculture.

separately, as given in table I. At the bottom of the table will be found the values for the probable error of the mean and the probable error of a single observation. These were calculated from the following formulas: P.E. mean $= \pm 0.6745\sqrt{\frac{\Sigma d^2}{n(n-1)}}$; P.E. sing. $= \pm 0.6745\sqrt{\frac{\Sigma d^2}{(n-1)}}$; where "n" is the number of variates (in this

TABLE I

COMPOSITION OF FIFTY-ONE ORANGES, WASHINGTON NAVEL VARIETY

Orange no.	Degrees brix	Percentage of sugar	Percentage of acid	Sol. sol. / acid ratio	Orange no.	Degrees brix	Percentage of sugar	Percentage of acid	Sol. sol. / acid ratio
1	12.80	9.63	0.98	13.05	30	13.70	10.91	1.07	12.80
2	13.10	10.30	0.98	13.35	31	14.00	10.93	0.96	14.60
3	12.50	9.46	1.08	11.60	32	13.70	10.68	1.14	12.00
4	13.70	10.44	1.14	12.00	33	15.30	11.92	1.15	13.30
5	14.40	11.17	1.06	13.60	34	13.85	10.83	1.06	13.05
6	15.00	11.14	1.06	14.15	35	13.20	10.31	1.02	12.95
7	13.90	10.85	0.84	16.55	36	14.70	11.51	0.86	17.10
8	13.40	10.43	0.98	13.70	37	14.75	10.96	1.04	14.20
9	13.70	10.94	0.93	14.75	38	15.30	11.57	1.29	11.85
10	13.70	10.65	0.84	16.30	39	15.30	11.46	1.23	12.45
11	13.55	10.71	0.90	15.05	40	13.75	10.88	0.91	15.10
12	13.35	10.14	1.15	11.60	41	13.40	10.18	1.29	10.40
13	13.20	10.35	0.94	14.05	42	13.35	10.33	1.19	11.20
14	13.95	10.85	0.98	14.25	43	13.45	10.09	1.24	10.85
15	14.30	10.83	0.96	14.90	44	14.45	11.00	0.94	15.35
16	15.05	11.59	0.95	15.85	45	12.80	10.05	1.15	11.15
17	14.90	11.80	1.02	14.60	46	14.00	10.76	1.27	11.00
18	13.20	10.30	1.09	12.10	47	14.70	11.26	0.86	17.10
19	15.25	12.05	1.00	15.25	48	14.90	11.44	0.98	15.20
20	13.40	10.53	1.02	13.15	49	12.60	11.35	1.07	11.80
21	14.85	11.18	1.11	13.40	50	14.80	10.88	1.31	11.30
22	13.40	11.35	1.01	13.25	51	14.10	11.13	1.16	12.15
23	14.45	11.49	0.82	17.60					
24	13.80	10.99	0.91	15.15	Mean	14.00	10.89	1.05	13.60
25	13.00	10.20	1.14	11.40					
26	14.45	11.28	1.22	11.85	P.E. mean	±0.07	±0.06	±0.01	±0.17
27	14.30	11.15	1.05	13.60					
28	14.60	11.61	1.12	13.05	P.E. sing.	±0.5	±0.4	±0.09	±1.3
29	14.55	11.40	0.87	16.70					

case fifty-one), and Σd^2 is the sum of the squares of the deviations of each measurement from the mean. For example, in the column under brix, table I, "d" is the deviation of 12.80 from 14.00, etc.

The probable error of a single sample and the probable error of the mean are connected in the following manner: P.E. mean $= \frac{\text{P.E. sing.}}{\sqrt{n}}$, so that after a value for P.E. sing. has been found, the value of P.E. mean for any desired number of fruits may be calculated by substituting this number for "n" in the formula. Thus if P.E. sing. has been found to be 0.5, P.E. mean for a sample of twenty-five fruits is $\frac{0.5}{\sqrt{25}}=0.1$.

The values in table II, giving the odds, may be utilized under the two following conditions. In the first place, it may be used in connection with the analytical results obtained from a single lot of fruit to estimate the degree of assurance that an accuracy between certain limits has been attained. For example, the average sugar content (in table I) was 10.89. If a second sample of fifty-one fruits had been taken at the same time and under the same conditions, we would probably not have obtained exactly this value.

TABLE II*

TABLE OF ODDS

Coefficient	Odds	Coefficient	Odds
1.0	1.00 to 1	3.4	44.87 to 1
1.5	2.21 to 1	3.6	64.79 to 1
2.0	4.64 to 1	3.8	95.15 to 1
2.5	9.89 to 1	4.0	142.26 to 1
2.8	15.95 to 1	4.2	215.92 to 1
3.0	22.26 to 1	4.4	332.33 to 1
3.2	31.36 to 1	4.6	519.83 to 1

* The values in this table were selected from a table by PEARL and MINER (6). Original article should be consulted for a complete list of values.

But the P.E. mean, ±0.06, indicates that the chances are even (1 to 1) that the value found would have been between 10.95 and 10.83. In addition to this information, table II shows that the chances are 9.89 to 1 that the value would have been between 10.89 plus (2.5×0.06) and 10.89 minus (2.5×0.06), that is, between 11.04 and 10.74.

Considering the probable error of a single sample in connection with table II, the P.E. sing. was found to be 0.4. This means that if one more fruit had been taken, the chances are even that its value would have been between 10.89+0.4, and 10.89−0.4. In other words, half the fruits in table I should have sugar values between 11.29 and 10.49, and half should be outside these limits. Table I shows that twenty-four oranges are within these limits and twenty-seven outside. Table II indicates further that the chances are 4.64 to 1 that no single sample would deviate from 10.89 by as much or more than 2.0 times 0.4: that is, of the fifty-one fruits in table I, about nine should be outside the limits 11.69 to 10.09, and forty-two should be within them. A count shows that in this case five are outside and forty-six within.

In the second place, table II may be applied in an entirely different case, namely, when comparing the analytical results from two different lots of fruit in order to estimate the degree of assurance that the difference shown between them is significant. For example, in table VII it is shown that the refractive index of the juice of the Eureka strain of lemons was 44.6±0.2, while that of the Shade Tree strain was 45.7±0.3. The difference is 1.1. What are the chances that this difference is significant and not due merely to a sampling error? This calculation is made from the following formula: $\frac{\text{difference}}{\text{P.E. of difference}} = \frac{1.1}{\sqrt{(0.2)^2+(0.3)^2}} = \frac{1.1}{0.36} = 3.0$. The figure 3.0 is here termed the coefficient of odds, and its value is sought in column 1 in table II, from which it appears that the odds are about 22 to 1 (judging from these data, at least) that the juice of lemons from the Shade Tree strain is higher with respect to refractive index. Table II applies only in those cases in which the difference between two results may be expected to occur in either direction. For a table showing odds when it is known that the difference between two results will be in one direction only, see Wood (**11**, p. 26).

Formulas for calculating number of fruits for sample

Two general sets of conditions may be recognized under which samples are collected for analysis: (1) When samples are taken from each of two or more different lots of fruit, with the object of later comparing them, to determine whether the differences between them are significant, and what the odds are that this is so. (2) When a sample is taken from a single lot of fruit for the purpose of obtaining a figure that will represent the composition of that lot, and to attain a certain assurance that this figure is correct within certain desired limits.

HAYNES and JUDD (3) have studied the requirements under the first condition. They proposed the following formula for use in calculating the number of individuals to include in a sample in order that a certain difference between two averages may be considered significant: $N=2\left(\frac{3\times p}{5}\right)^2$. N is the "number of samples which must be taken in order that there may be a probability of 0.957[2] that a 5 per cent difference is significant"; 3 is the coefficient in the "table of odds" (table II), and thus is equivalent to odds of 22 to 1; "p" is the probable error of a single sample and must be determined experimentally (in this case by analyzing individual fruits).

Other values than 3 and 5 may be assumed to meet the conditions of the experiment; therefore, in order to make comparisons with what is to follow, it is desired to express the preceding formula in general terms as follows: $N=2\left(\frac{\text{coefficient of odds}\times\text{P.E. sing.}}{\text{difference}}\right)^2$ (*formula 1*). To illustrate the use of this formula, data may be taken from HAYNES and JUDD's paper. Working with apples, they found the mean titration value to be 10.20 with a P.E. sing. of 0.78, and the latter is thus 7.7 per cent of the mean. To get an assurance of 30 to 1 that a 5 per cent difference is significant: $N=2\left(\frac{3.2\times 7.7}{5}\right)^2=49$ apples.

The problem under the second condition may now be considered. We wish a general formula that will connect the number in the sample with the probable error of a single fruit and with the coefficients in the "table of odds" (table II). In table I it was shown that the mean sugar content was 1089.$\pm$0.06. What are the chances that the "true" value is within the limits $\pm$0.17? The chances are found in the following way (MERRIMAN 5): $\frac{0.17}{0.06}=2.8$, and looking up the coefficient 2.8 in table II, we find the chances are about 16 to 1 that the error in 10.89 is not more than $\pm$0.17.

[2] The expression 0.957 may be thought of as indicating a probability of 957 out of 1000, which represents a ratio of 957 to 43, or about 22 to 1.

This relation may now be expressed in general terms by putting "deviation" for ± 0.17, where it is to be the deviation above or below the mean, which we wish to use as a limit for accuracy; then putting "P.E. mean" for 0.06, and "coefficient of odds" for 2.8, we have: $\frac{\text{deviation}}{\text{P.E. mean}} = \text{coefficient of odds}$, but $\text{P.E. mean} = \frac{\text{P.E. sing.}}{\sqrt{N}}$ (Wood 11), and substituting this value, the equation becomes $\frac{\text{deviation}}{\frac{\text{P.E. sing.}}{\sqrt{N}}} = \text{coefficient of odds}$, from which

$$N = \left(\frac{\text{coefficient of odds} \times \text{P.E. sing.}}{\text{deviation}}\right)^2 \quad (\textit{formula 2}).$$

In illustration of the use of this formula, table VI shows that fifty grapefruits from tree no. 1 had an average brix of 13.15 and the P.E. sing. was 0.35. What number of fruits are required to give odds of 10 to 1 that the brix of that number will be correct to ± 0.15? Table II shows that for odds of 10 to 1, the coefficient of odds is 2.5, therefore $N = \left(\frac{2.5 \times 0.35}{0.15}\right)^2$ = thirty-four grapefruits. No account is taken of errors in the method of analysis, since in the present case analytical errors are small as compared with the variability of the individual fruits with respect to the constituent. If it is desired to take analytical errors into account also, see Waynick (10) and Robinson and Lloyd (7).

Comparison of formulas

Although formulas 1 and 2 appear to be very similar, the first in fact giving values just double those of the second, certain essential differences should be pointed out. Formula 1 applies when *two different lots* are being compared, in which case the significance of the difference between them is affected by the sampling error of each lot. Formula 2 applies to the analytical results of a *single lot* only, its own error being the only one involved. Such a condition arises when an analysis is made for the purpose of reporting the composition of a product with respect to a certain constituent, or when an analysis is made to determine whether a constituent has reached a certain required value.

Accuracy of formulas

In the preceding paragraphs it was found that the use of formulas 1 and 2 gave forty-nine fruits as the required number in one illustrative case, and thirty-four as the required number under the other set of conditions. We should not be justified, however, in concluding from this test that forty-seven would be too few in the first case, and thirty-six would be more than enough in the second. With either formula it is seen that the number N depends for its value upon the value of the probable error of a single sample, and therefore it becomes necessary to inquire how variable this value is, and what effect changes in its value have upon N.

TABLE III

DIFFERENT VALUES OBTAINABLE FROM SAME LOT OF FRUIT

Calculations after the following number of fruits analyzed	Taken in order of analysis		Taken in order rearranged by lot			
			First rearrangement		Second rearrangement	
	P.E. sing. found	No. of fruits required	P.E. sing. found	No. of fruits required	P.E. sing. found	No. of fruits required
10	1.1	22	1.4	36	1.0	18
15	1.1	22	1.4	36	1.2	26
20	1.0	18	1.4	36	1.2	26
25	1.1	22	1.4	36	1.2	26
30	1.1	22	1.4	36	1.4	36
35	1.1	22	1.4	36	1.4	36
40	1.1	22	1.4	36	1.3	31
45	1.2	26	1.3	31	1.3	31
51	1.3	31	1.3	31	1.3	31

It is instructive to note what values would have been obtained if the value of P.E. sing. had been taken, not after fifty-one fruits had been analyzed, but after the analysis of say ten fruits, or after fifteen, or twenty-five. The different values for P.E. sing. and N that were obtainable in this manner calculated from formula 1 are shown in table III. It is thus found the P.E. sing. varied from 1.0 to 1.3, which values, substituted in the formula, caused the value of N to vary from 18 to 31. Formula 2 would likewise have given variable values, but the actual figures would have been one-half as large.

The fruits in table I were analyzed in the order of size, number one being the largest. It may be urged that therefore we do not have a true random sample, or that there is a correlation between size and composition. The correlation coefficient between size and the soluble-solids-acid ratio, however, was calculated by the method recommended by TOLLEY (9), and was found to be 0.158, with a probable error of 0.092, which does not indicate any significant correlation.

In order to partially eliminate the size of the fruit as a factor, the order in table I was rearranged by lot. With the new order,

TABLE IV

RESULTS OF CALCULATIONS OF PROBABLE ERROR BASED ON ANALYSIS OF GROUPS OF TEN FRUITS EACH

GROUPS OF 10 FRUITS EACH	SOLIDS-ACID RATIO	
	P.E. sing.	No. required for desired assurance
Group 1	1.0	18
2	0.9	15
3	1.3	31
4	1.0	18
5	1.6	47
6	1.4	36
7	1.3	31
8	1.3	31
9	1.3	31
10	0.5	5
11	1.8	59
12	1.3	31
13	1.1	22
14	0.9	15
15	1.2	26

the values of P.E. sing. and N were calculated after ten fruits were analyzed, after fifteen, etc. The results are shown in table III. P.E. sing. was found to vary from 1.3 to 1.4, causing N to vary from 31 to 36. Another rearrangement by lot is shown in the last two columns of table III. Values of P.E. sing. vary from 1.0 to 1.4, causing N to change from 18 to 36. In both these cases, values by formula 2 would also have been variable, but of course would have been just half as large numerically.

Use of small numbers to calculate probable error of single fruit

It may be inquired what the P.E. sing. would have been for different lots of ten fruits each. Groups of ten each were selected by lot and the values of P.E. sing. and N calculated. Strictly speaking, when the number involved is small, say ten, the formula for P.E. only gives approximate results (BRUNT 1). The value of P.E. sing. for the ratio is thus shown to vary from 0.5 in group 10 to 1.8 in group 11, causing a change in N from 5 to 59 (table IV). One trial with a small number of fruits would not be adequate for the determination of the value of P.E. sing. and of N, at least with such variable material as oranges.

Probable error of a probable error

The preceding discussion indicates that variable values were found for N, depending on the value found for P.E. sing. To obtain an idea of the variability of P.E. sing. and of N in the manner described (that is, by obtaining the results given by several different groups containing different numbers) is tedious and unsatisfactory. A more convenient method of judging the accuracy of P.E. sing. and N is desired. It is plain that the probable error calculated from the analysis of fifty fruits is more representative of the lot than that calculated from ten fruits. The relation of the error in the probable error to the number of fruits analyzed is given by the expression (BRUNT 1, p. 57): Probable error of P.E. sing. = P.E. sing. $\times \frac{0.4769}{\sqrt{n-1}}$ (*formula 3*). Thus if 1.3 is the P.E. sing. for the soluble-solids-acid ratio (table I), then the probable error of $1.3 = 1.3 \times \frac{0.4769}{\sqrt{51-1}} = 0.09$, or about 0.1. In other words, the "true" value of P.E. sing. is probably between 1.2 and 1.4. We may obtain an estimate of the limits of N by substituting 1.2 and 1.4 successively in the formulas; in this case N is found to be 26 or 36 for formula 1, and 13 or 18 for formula 2.

Ordinarily it will be sufficient to consider the probable limits of the value of N by approximations made by the use of formula 3 in the manner indicated. If it is found desirable to do so, however, a formula may be used for the correction. If we rearrange formula 1 to read: $N = 2\left(\frac{\text{coefficient}}{\text{difference}}\right)^2 (\text{P.E. sing.})^2$, and apply the method described by GOODWIN (2), we find that deviation produced in the value of N by an error in the value of P.E. sing. is as follows:

$$\text{Deviation in N} = \frac{d\left[2\left(\frac{\text{coefficient}}{\text{difference}}\right)^2(\text{P.E. sing.})^2\right]}{d(\text{P.E. sing.})} \times \text{error in P.E. sing.}, = 4\left(\frac{\text{coefficient}}{\text{difference}}\right)^2 \times \text{P.E. sing.} \times \text{error in P.E. sing.}$$ (*formula 4*).

To apply this formula to a particular case, we find from table III that the P.E. sing. for fifteen fruits was 1.1; the error in 1.1 is found by substituting in formula 3 to be $1.1 \times \frac{0.4769}{\sqrt{15-1}} = 0.14$. If we wish odds of 22 to 1 for a difference of 1.0 in ratio, we obtain, by substitution in formula 4: Deviation in $N = 4 \times \left(\frac{3.0}{1.0}\right)^2 \times 1.1 \times 0.14 =$ six fruits, therefore the corresponding value, 22, found in table III, is in error by six fruits, and the probable number extends from 16 to 28.

The corresponding formula for applying a correction to formula 2 is: $\text{Deviation in N} = 2 \times \left(\frac{\text{coefficient}}{\text{difference}}\right)^2 \times \text{P.E. sing.} \times \text{deviation in P.E. sing.}$ (*formula 5*).

Data on other lots of oranges

The discussion thus far has related to the data from only one lot of oranges from a single tree. Fruits from four other trees were obtained and analyzed in the same manner. The number of fruits used was small, but some idea of the accuracy of the probable errors can be obtained by applying formula 3. The data are shown in table V, and serve to indicate values of P.E. sing. that may be expected in dealing with different lots of oranges.

Data on grapefruit

Fifty fruits were taken at random from a grapefruit tree in one grove, and a corresponding number from another tree located in another grove. The fruits were analyzed individually and the mean and P.E. sing. determined. To save space, the complete analyses are not given, but the results are summarized in table VI. From this table it is seen that P.E. sing. of the fruit from the two lots is approximately the same with respect to brix and sugar, but P.E. sing. for acid and for ratio is considerably different in the two lots.

Data on lemons

That different lots of fruit show different values for P.E. sing. is also apparent from the analysis of individual lemons. In table VII will be found the results of the analysis of thirty lemons from two different lemon trees, each tree representing a different strain

TABLE V

SHOWING DIFFERENT VALUES OF P.E. SING. WITH DIFFERENT LOTS OF ORANGES

TREE NO.	NO. OF FRUITS IN SAMPLE	DEGREES BRIX		PERCENTAGE ACID		$\frac{\text{Sol. sol.}}{\text{acid}}$ ratio	
		Mean	P.E. sing.	Mean	P.E. sing.	Mean	P.E. sing.
2	12	13.70	0.5	0.87	0.07	15.9	1.1
3	13	15.00	0.5	0.86	0.04	17.4	0.8
4	12	11.80	0.4	0.79	0.08	15.2	1.3
5	9	12.45	0.3	1.46	0.10	8.6	0.7

TABLE VI

COMPARISON OF COMPOSITION OF FRUIT FROM TWO GRAPEFRUIT TREES

TREE NO.	TOTAL NO. OF FRUITS	BRIX		PERCENTAGE SUGAR		PERCENTAGE ACID		SOLIDS-ACID RATIO	
		Mean	P.E. sing.	Mean	P.E. sing.	Mean	P.E. sing.	Mean	P.E. sing.
1	50	13.15	0.35	8.16	0.27	2.29	0.01	5.8	0.2
2	50	12.30	0.35	7.89	0.29	1.65	0.09	7.5	0.4

of the Eureka variety. While too much reliance cannot be placed on the values obtained by analyzing fifteen fruits, it is seen from the table that the two lots of fruit probably have different values of P.E. sing. with respect to three of the characters of which analytical results were obtained.

Further precautions regarding use of formulas

Two further precautions may now be added regarding the use of the formulas. When the value of P.E. sing. has been found for one tree or lot of fruit, it must not be assumed that another tree or lot will have the same value (compare acidity of two grapefruit trees, table VI). When two trees or lots of fruit are found to have the same value for P.E. sing with respect to one constituent, it must not be assumed that they agree also with respect to other constituents (compare trees no. 1 and no. 2, table VI, with respect to brix and ratio).

TABLE VII

VARIABILITY IN COMPOSITION OF INDIVIDUAL LEMONS, EUREKA VARIETY

Eureka strain*					Shade tree strain*				
Lemon no.	Sp. Gr. of fruit	Percentage rind	Refrac. index of juice	Acidity of juice cc. NaOH	Lemon no.	Sp. Gr. of fruit	Percentage rind	Refrac. index of juice	Acidity of juice cc. NaOH
1........	0.92	40	42.9	27.2	1........	0.96	41	42.1	24.3
2........	0.96	40	44.2	28.5	2........	0.94	59	45.4	24.4
3........	0.94	50	44.2	28.8	3........	0.96	50	45.8	24.9
4........	0.96	49	44.8	29.7	4........	0.98	36	44.9	26.8
5........	0.95	49	45.2	27.9	5........	0.98	47	46.8	21.8
6........	0.94	46	44.6	30.7	6........	0.95	62	47.0	24.0
7........	0.96	48	45.6	30.5	7........	0.95	56	45.8	24.5
8........	0.95	35	44.0	28.7	8........	0.96	54	45.6	22.9
9........	0.94	46	43.6	28.1	9........	0.98	47	45.4	22.0
10........	0.96	50	46.3	25.1	10........	0.99	54	47.6	20.8
11........	0.95	50	45.0	27.8	11........	0.96	48	45.7	26.6
12........	0.95	48	45.6	28.0	12........	0.98	39	45.7	26.3
13........	0.96	57	43.8	29.3	13........	0.98	51	46.2	22.6
14........	0.96	39	43.9	28.0	14........	0.98	54	......	20.9
15........	0.97	49	45.4	28.7	15........	0.97	59	......	22.0
Mean....	0.95	46	44.6	28.5	Mean.....	0.97	50	45.7	23.7
P.E. mean	±0.002	±1.0	±0.2	±0.2	P.E. mean.	±0.003	±1.3	±0.3	±0.4
P.E. sing..	±0.007	±4.0	±0.6	±0.9	P.E. sing..	±0.010	±5.0	±0.9	±1.3

* Strains described by SHAMEL, SCOTT, and POMEROY (8).

Comparison of standard formula with Peter's formula for calculating probable error of single observation

Two general methods for calculating the value of P.E. sing. are as follows:

STANDARD FORMULA

$$\text{P.E. sing.} = \pm 0.6745 \sqrt{\frac{\Sigma d^2}{n-1}}$$

PETER'S FORMULA

$$\text{P.E. sing.} = \pm 0.8453 \frac{\Sigma d}{\sqrt{n(n-1)}}$$

Thus, to use the standard formula, the sum of the squares of the deviations must be found, while with Peter's formula only the sum of the deviations (taken without regard to sign) is needed. Inasmuch as the latter method is more convenient, it seemed profitable to show the difference in the value of P.E. sing. given by the two methods. In table VIII are shown the comparative values found.[3] It is seen that the difference in the value of P.E. sing. by the two methods is at least not more than is shown between two groups of even the same lot of fruit. Hence no large error would have been introduced by the use of the more convenient Peter's formula.

TABLE VIII

COMPARISON OF STANDARD FORMULA WITH PETER'S FORMULA FOR CALCULATING PROBABLE ERROR OF SINGLE OBSERVATION

No. of fruits in sample	P.E. sing. observation: Solids-acid ratio		No. of fruits in sample	P.E. sing. observation: Percentage sugar	
	Standard formula	Peter's formula		Standard formula	Peter's formula
10	1.09	1.08	10	0.39	0.40
15	1.01	.99	15	0.34	0.34
25	1.09	1.10	20	0.44	0.42
30	1.10	1.13	25	0.42	0.43
40	1.09	1.02	35	0.40	0.40
45	1.20	1.22	45	0.40	0.40
51	1.26	1.29	51	0.39	0.38

Summary

1. Formulas are given, for use under two different conditions of sampling, to determine the number of fruits required in a sample in order to give a desired assurance that a certain accuracy has been attained.

2. Approximately 250 fruits of oranges, lemons, and grapefruit were analyzed individually, and the probable errors calculated. The data so obtained were applied to the formulas, and numerical examples worked out to illustrate their use.

3. It is shown that the values given by the formulas are only approximately correct. The sources of error are discussed, and formulas given by which the amount of this inaccuracy may be estimated under different conditions.

[3] Computations are made much easier by the use of tables given by MELLOR (4).

4. Analyses of fruits taken from different orange, lemon, and grapefruit trees are given, showing the variability of the fruits of different trees with respect to brix of juice, percentage of sugar, acidity, etc., and the values of the probable errors that such variability produced.

The writer wishes to express appreciation to Mr. E. M. CHACE and Mr. C. G. CHURCH for cooperation in obtaining the analytical data and for criticism of the manuscript.

UNITED STATES DEPARTMENT OF AGRICULTURE
LABORATORY OF FRUIT AND VEGETABLE CHEMISTRY
LOS ANGELES, CAL.

LITERATURE CITED

1. BRUNT, D., The combination of observations. Cambridge. 1917.
2. GOODWIN, H. M., Precision of measurement and graphical methods. New York. 1919.
3. HAYNES, D., and JUDD, H. M., The effect of methods of extraction on the composition of expressed apple juice, and a determination of the sampling error of such juice. Biochem. Jour. **13**: 272–277. 1919.
4. MELLOR, J. W., Higher mathematics for students of chemistry and physics. New ed. London. 1919.
5. MERRIMAN, M., A text-book on the method of least squares. New York. 1913.
6. PEARL, R., and MINER, J. R., A table for estimating the probable significance of statistical constants. Maine Agric. Exp. Sta. Bull. 226. 1914.
7. ROBINSON, G. W., and LLOYD, W. E., On the probable error of sampling in soil surveys. Jour. Agric. Sci. **7**: 114–153. 1915–16.
8. SHAMEL, A. D., SCOTT, L. B., POMEROY, C. S., and DYER, C. L., Citrus-fruit improvement: A study of bud variation in the Eureka lemon. U. S. Dept. Agric. Bull. 813. 1920.
9. TOLLEY, H. R., The theory of correlation as applied to farm-survey data on fattening baby beef. U. S. Dept. Agric. Bull. 504. 1917.
10. WAYNICK, D. D., Variability in soils and its significance to past and future soil investigations. I. A statistical study of nitrification in soils. Univ. Calif. Pub. Agric. Sci. **3**: 243–270. 1918.
11. WOOD, T. B., The interpretation of experimental results. Jour. Bd. Agric. (London) Sup. **7**: 15–37. 1911.